INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

# A Proficient Method for Understanding of Indian Sign Language using Machine Learning

[1]KESANAPALLI NIRIKSHANABABU, [2]A. CHENNAKESAVA REDDY

[1]PG Scholar, Dept. of MCA, Newton's Institute of Engineering, Guntur, (A.P)

[2]Assistant professor, Dept. of CSE, Newton's Institute of Engineering, Guntur, (A.P)

***Abstract**:* Sign Language is used for non-verbal communication. People who have trouble hearing or speaking utilise sign language as a means of communication. However, most people have trouble deciphering the hand motions of the disabled because they are unfamiliar with sign language. A translator is often required for two parties if one party has speech or hearing impairments and the other party does not. In this paper, we propose a system that can convert the hand gestures for the numbers 0-9, the English alphabet, and a few English words from Indian Sign Language (ISL) into comprehensible text, and vice versa, to help people with special needs communicate more effectively with the people around them. Machine Learning and image processing methods are used for this purpose.

Several neural network classifiers for gesture recognition are created, evaluated, and ranked according to their effectiveness.

***Keywords**: I*ndian Sign Language, hand gestures, interpreter, SURF, Convolutional Neural Network, Recurrent Neural Network, K-means clustering, Support Vector Machine.

## I INTRODUCTION

Different cultures use a wide range of sign languages. About 300 distinct sign languages are in use in different regions of the globe today. This is due to the fact that sign languages were independently established by persons of various racial and cultural backgrounds.

Perhaps there is no universally accepted kind of sign language in India. There are lexical differences and regional variants of Indian Sign Language spoken in various areas of the country. The Indian Sign Language (ISL) has been standardised in recent years.

There are two main types of hand gestures in ISL: (i) stationary gestures and (ii) moving gestures. Figure 1 depicts the static ISL hand motions for the numerals

0-9, the English alphabet (A-Z), and a few English phrases.

About 50 million individuals in India have some kind of speech or hearing impairment, according to the latest census data (2011). However, in India, there are fewer than 300 qualified sign language translators. Because of this, those who have trouble communicating typically. lonely and solitary since they are unable to communicate with regular people. Their personal and professional lives are profoundly affected by this.

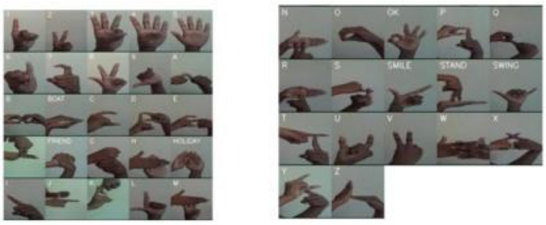


Fig. 1. ISL Hand Gestures

In light of these difficulties, this work proposes a real-time automated system that can translate English words into International Sign Language (ISL). Those with communication difficulties may now do so with ease thanks to this technology. This may inspire individuals to improve their skills and see they have room to grow as people. The suggested system serves dual purposes: (i) translating hand movements into text, and (ii) translating spoken commands into gestures. Neural network classifiers perform the transformation from gesture to text.

The Google Speech Recognition API is used to translate spoken commands into hand gestures.

This work is concerned with the most precise possible translation between Indian Sign Language and English, and between English spoken words and Indian Sign Language gestures. To do this, many neural network classifiers are created and evaluated for their efficacy in recognising hand gestures. The best performing classifier is selected for use in creating an ISL gesture-to-English text and speech-to-ISL gesture-to-English text translation app.

## II **LITERATURE SURVEY**

The system introduced by Feng-Sheng Chen et al. (2003) tracks the moving hand and analyses the hand-shape variation and motion information as the input to the HMM based recognition system, allowing it to recognise "dynamic gesture," in which a gesture was performed singly in a complex background using 2D video input. They report that 4-state HMM generates the greatest results in their experiments when it comes to simulating the gesture.

Coding has been put in place either for contour information exclusively or a mix of contour and motion information, and a manual region extraction technique is used for each input picture sequence. In order to use the retrieved data for training and recognition, the vector sequences are transformed into symbol sequences by a quantization procedure. Each of the 20 various motions has on average 60 picture sequences gathered for it during the training phase, and another 1200 sequences are collected during the testing phase. Method 1 has a 97% recognition rate when testing with training data and a 90.5% recognition rate when testing with testing data, whereas Method 2 has a 98.5% recognition rate when testing with training data and a 93.5% recognition rate when testing with testing data.

Detecting the existence of human hands inside a picture and categorising the hand form is a challenging problem, but researchers Eng-Jon Ong and Richard Bowden (2004) provided a unique, unsupervised way to train an efficient and robust detector. Their method involves identifying the position of the hands inside a greyscale picture using a boosted cascade of classifiers. So, we used the k-medoid clustering method with a distance metric based on form context to group together our picture library of hands into groups of hands that are visually similar to one another. The unsupervised clustering approach produced sets of hand forms, and from these, a tree of enhanced hand detectors was constructed, with the top layer dedicated to generic hand detection and the second layer specialising in categorising these shapes. With an unseen library of 2509 photos, the detector achieved a 99.8% success rate, while the shape classifier had a success rate of 97.4%.

View-specific hand posture identification using an object recognition approach recently presented by Viola and James was the subject of research by Kolsch, M., and Turk, M. (2004).

To begin, they proved that the integral picture method was adequate for the problem of recognising hands. The necessity for computationally demanding training was then removed with the introduction of the qualitative measure that can be thought of as an a priori estimate of detectability. Finally, detection technique settings were optimised, leading to significant gains in both speed and accuracy. They argued that the rectangular feature-classification approach was most

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

suited for detecting convex appearances with internal grey-level changes.

Henrik Jonsson (2008) has conducted research on the potential for developing an automated system to detect Swedish sign language from visual data. The project's concentration on segmenting hands under challenging lighting settings meant that the purpose of monitoring motions was not fulfilled. To do this, we had to learn about the theory behind visual processing and document it, all while putting it into practise in software. The form of a hand may be extracted automatically using a skin colour model under particular conditions.

## III. PROPOSED METHODOLOGY

The aforementioned section outlined the two primary functions of the proposed system for ISL interpretation: Both (i) gesture-to-text and (ii) speech-to-gesture translation are possible.

First, a Symbolic Language Translator

There are four main stages in the gesture-to-text conversion process: Dataset collection (i), segmentation (ii), feature extraction (iii), and classification (iv) are the four main steps. Figure 2 shows a conceptual model of the gesture-to-text conversion process. The initial phase of gesture-to-text technology the groundwork for a multimedia, multilingual Indian Sign Language dictionary was laid by Tirthankara Dasgupta et al. in 2008. They demonstrated a multimedia, cross-platform, Indian Sign Language (ISL) dictionary-making application. absence of ISL expertise and the absence of available instructional tools make ISL a linguistically under-investigated language with no source of well-documented electronic data. Signs relating to a particular text may be assigned using the suggested approach. Using the Hammons's framework, the present approach makes it easier to annotate Indian signs with phonological information. The Hammons's string that is created may be sent into an avatar module to make a movable sign.

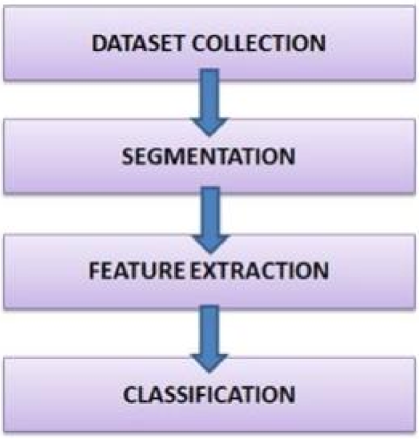


Fig. 2. Concept Diagram - Gesture to Text Conversion

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

Dataset collection is the conversion. Images representing ISL hand motions for the numerals 0-8, the alphabet, and certain English words are gathered to form a dataset. Once the dataset is complete, all of the included photos are pre-processed to eliminate noise and obscure any distracting features.

As a result, the system's efficiency, accuracy, and performance are all enhanced by doing pre-processing on the pictures before feeding them to a classifier. This is a crucial stage in the picture categorization process.

The steps below outline the pre-processing of data.

First, making all of the pictures the same size.

2) Changing a colour picture into a black-and-white one

Third, the Blurring of the Median

4) Covering/Uncovering Your Face/Body

5) Sharp-edge detection (using a Canny edge detector)

Feature extraction is the next stage in the process of translating gestures into text.

The processed photos are then used for feature extraction.

In computer vision and picture categorization, feature extraction plays a crucial role. To prepare the data for processing by the classification algorithm, the raw data (pictures) must be converted into numerical characteristics. Data integrity is maintained even after the picture is transformed to numbers.

The SURF technique is used here for feature extraction. The SURF may either be used to identify features or to describe them. Object identification, picture categorization, and similar tasks often make use of it. This approach for representing and comparing photos is both quick and reliable. In a picture, it may help identify blobs. In order to compute the SURF features, we use the determinants of Hessian matrices to locate the spots of interest in the picture that have the relevant characteristics. Scale invariant descriptors are built for every point of interest identified in the preceding step.

Inequalities (1) and (2) provide the Hessian matrix and its determinant.

$$H(f(x,y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \quad (1)$$

$$det(H) = D_{xx}D_{yy} - (0.9D_{xy})^2 \quad (2)$$

Machine Learning methods such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Support Vector Machines (SVMs) use the retrieved picture attributes as input.

A supervised machine learning approach, support vector machines employ a hyper-plane to categorise input. K-means clustering classifier and the Bag of Visual words (BoV) model are used in conjunction with the SVM classifier to improve accuracy.

K-means clustering is a kind of unsupervised classification that divides a dataset into as many clusters as there are classes (where k is the total number of categories).

The k-means clustering model's output is used as input to the BoV model, which performs the classification. Using the number of "visual words" (image characteristics) present in a picture, the BoV model further categorises the images. The SVM receives the results from the BoV classifier. The SVM is the primary classifier used in both development and evaluation.

About 80% of the dataset is utilised for training in the SVM classifier, while the remaining 20% is used for testing.

Models of CNN and RNN classifiers are developed, and their capabilities in gesture recognition are evaluated. The dataset has three sections, one each for CNN and RNN:

Sixty percent of the data is utilised for training, twenty percent for testing, and the last twenty percent for validation.

The best image classifier for gesture recognition is determined by comparing the performance of the aforementioned classifiers.

The most accurate classifier is selected for ISL gesture detection in real-time video. It's just as easy to recognise motions in real-time video as it is in a still picture. Here are the procedures required to detect gestures in streaming video in real time.

1) A webcam is used to record the footage.

2) The video is taken as still images at each frame.

Third, the obtained pictures are reduced in size and pre-processed.

Extraction of SURF Features, Step 4

Fifth, the picture characteristics are sent to a classifier.

6) Anticipated gestures

B. Conversational Gesture Interpretation

The following steps are used to translate vocalisations into ISL hand motions:

Two steps are involved: 1) Speech-to-text conversion, and 2) database comparison of the resulting text.

ISL Gesture Output Displaying the Corresponding

Audio and the Google Speech Recognition API are used to convert speech to text. Figure 3 is a conceptual design for translating spoken language into hand gestures.

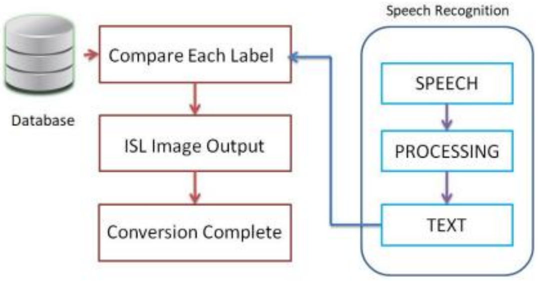


Fig. 3. Concept Diagram - Speech to Gesture Conversion

## IV EXPERIMENTAL RESULTS

First, a Symbolic Language Translator

The data set (refer to Fig. 1) includes ISL hand gestures for the digits 0 through 9, the letters A through Z, and seven English words: BOAT, FRIEND, HOLIDAY, OK, SWING, SMILE, and STAND. Therefore, there were 42 different types of photos in the dataset (9 + 26 + 7). Each picture category in the collection has 1200 unique photos.

As was indicated before, noise was removed and unwanted regions were masked in the photos included in the dataset. Figure 3 depicts the different picture pre-processing stages applied to a representative image from the dataset.

After the images in the dataset were pre-processed, the SURF feature matrix was computed for each one. In Fig. 4, we can see the SURF characteristics retrieved from an example picture. In Fig. 4, the SURF Feature points are represented by the blue circles of varied sizes. Different neural network classifiers were trained using the retrieved SURF features from all the photos, which were then saved in a pickle file. Below, we'll talk about how effective each of the evaluated classifiers was.

IJTE INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING

ISSN: 2057-5688

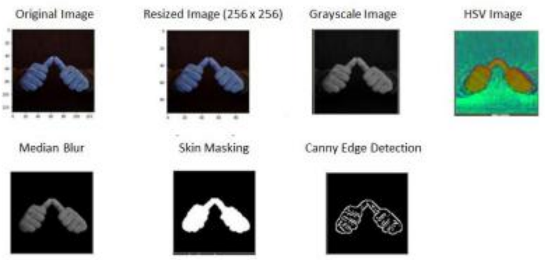
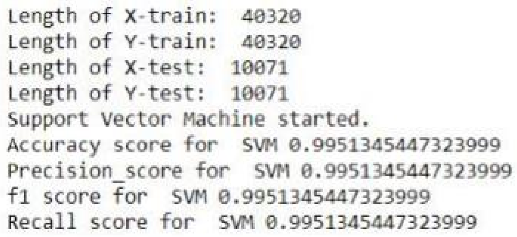


Fig. 4. Data pre-processing

Fig. 5. SURF Features

Before being sent to the SVM classifier, the input photos went via K-means clustering and a Bag of Visual Words classifier. Since the dataset contains 42 different types of photos, we set the k parameter for the Kmeans clustering classifier to 42. After using the k-means clustering technique, the visual words were gathered for both the test and train datasets. The dataset contains a total of 50391 pictures. The SVM model was trained using 40320 of these photos. The remaining 10071 photos were put through the classifier's paces for evaluation. The accuracy of the tests was around 99.5%. Aside from the accuracy score, the F1 score, and the recall rate were also computed. Figure 6 depicts them.



```
Length of X-train:  40320
Length of Y-train:  40320
Length of X-test:  10071
Length of Y-test:  10071
Support Vector Machine started.
Accuracy score for  SVM 0.9951345447323999
Precision_score for  SVM 0.9951345447323999
f1 score for  SVM 0.9951345447323999
Recall score for  SVM 0.9951345447323999
```

Fig. 6. Accuracy of SVM Classifier

Modelling and developing a Convolutional Neural Network in Python using the Kera's package. Roughly 30,240 photos were utilised to train the classifier model (representing roughly 60% of the dataset's total). Various training epoch lengths were used to improve the classifier's performance. The highest average accuracy achieved throughout testing was 88.89 percent.

Thirdly, we have a recurrent neural network that was built with the help of the Kera's package in Python. The classifier model was trained with around 30,240 pictures. Various training epoch lengths were used to improve the classifier's performance. As a whole, the testing accuracy peaked at roughly 82.3%.

The aforementioned findings suggest that the best accuracy in recognising hand gestures is achieved by combining K-Means clustering, BoV, and SVM classifiers.

As a result, it's a safer bet for recognising gestures.

B. Real-Time Video Recognition of Gestures

The SVM classifier was used to create a system that could recognise gestures in real time. ISL hand gestures may be translated into English by holding them up to a camera. In real-time video, it takes roughly 0.04 seconds to anticipate the hand motion. As can be seen in Fig. 7, the real-time gesture recognition system captures images for analysis.

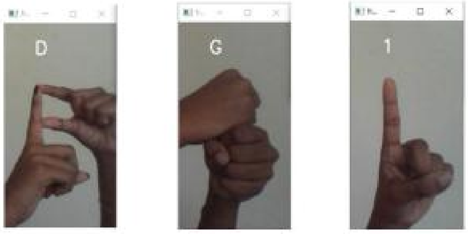


Fig. 7. Real-time gesture recognition

## V CONCLUSION:

As can be seen from the findings, the SVM classifier, in conjunction with the K-means clustering and BoV classifiers, performs the best when used to gesture recognition. Using the most effective SVM classifier (for gesture to text conversion) and the Google Speech Recognition API (for speech to gesture conversion), a user-friendly application that can read Indian Sign Language has been built. As a result, a more solid method for interpreting sign language has been created.

## ACKNOWLEDGMENT

We would like to thank PSG College of Technology, ECE Department for providing an opportunity to work on the project. We thank Dr. V. Krishnan, Professor (CAS) & Head In-charge, Department of Electronics and Communication Engineering for the encouragement and support that she extended towards this project work. We also thank our Programme Coordinator Dr. M. Santhanalakshmi, Associate Professor, Department of ECE and our Tutor Ms. P. Parvathi, Assistant Professor, Department of ECE for the advice and constructive feedback regarding various aspects of the project

## REFERENCES

[1] Cheok Ming Jin, Zaid Omar, Mohamed Hisham Jaward, ”A Mobile Ap plication of American Sign Language Translation via Image Processing Algorithms”, in IEEE Region 10 Symposium, on IEE Explore, 2016.

[2] Mr. Sanket Kadam, Mr. Aakash Ghodke Prof. Sumitra Sadhukhan, ”Hand

Gesture Recognition Software based on ISL", IEEE Xplore, 20 June 2019.

[3] Kartik Shenoy, Tejas Dastane, Varun Rao, Devendra Vyavaharkar, " Real-time Indian Sign Language (ISL) Recognition", IEEE Xplore: 18 October 2018.

[4] T Raghuveer, R Deepthi, R Mangalashri and R Akshaya, "A depth based ISL Recognition using Microsoft Kinect", ScienceDirect, 2018.

[5] Muthu Mariappan H, Dr Gomathi V, "Real time Recognition of ISL", IEEE Xplore: 10 October 2019.

[6] G. Ananth Rao a, P.V.V. Kishore, "Selfie video based continuous Indian sign language recognition system", ScienceDirect, 2018.

[7] G. Ananth Rao a, P.V.V. Kishore, "Sign Language Recognition Based on Hand and Body Skeletal Data", IEEE 2018.

[8] Suharjitoa, Ricky Anderson, Fanny Wiryanab, Meita Chandra Ariestab, Gede Putra Kusuma, "Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output", ScienceDirect, 2017.

[9] Prasadu Peddi (2023), Using a Wide Range of Residuals Densely, a Deep Learning Approach to the Detection of Abnormal Driving Behaviour in Videos, ADVANCED INFORMATION TECHNOLOGY JOURNAL, ISSN 1879-8136, volume XV, issue II, pp 11-18.

[10] Afreen Bari, Dr. Prasadu Peddi. (2021). Review and Analysis Load Balancing Machine Learning Approach for Cloud Computing Environment.Annals of the Romanian Society for Cell Biology,25(2), 1189–1195.

[11] Prasadu Peddi (2017) "Design of Simulators for Job Group Resource Allocation Scheduling In Grid and Cloud Computing Environments", ISSN: 2319-8753 volume 6 issue 8 pp: 17805-17811.